THE CYCLE STRUCTURE OF A MARKOFF AUTOMORPHISM OVER FINITE FIELDS

ALOIS CERBU, ELIJAH GUNTHER, MICHAEL MAGEE, LUKE PEILEN

Abstract. We begin an investigation of the action of pseudo-Anosov elements of $\mathrm{Out}(\mathbf{F}_2)$ on the Markoff-type varieties

$$\mathbb{X}_\kappa : x^2 + y^2 + z^2 = xyz + 2 + \kappa$$

over finite fields $\mathbb{F}_p$ with p prime. We first make a precise conjecture about the permutation group generated by $\mathrm{Out}(\mathbf{F}_2)$ on $\mathbb{X}_{-2}(\mathbb{F}_p)$ that shows there is no obstruction at the level of the permutation group to a pseudo-Anosov acting 'generically'. We prove that this conjecture is sharp. We show that for a fixed pseudo-Anosov $g \in \mathrm{Out}(\mathbf{F}_2)$, there is always an orbit of g of length $\geq C \log p + O(1)$ on $\mathbb{X}_\kappa(\mathbb{F}_p)$ where $C > 0$ is given in terms of the eigenvalues of g viewed as an element of $\mathrm{GL}_2(\mathbf{Z})$. This improves on a result of Silverman from [26] that applies to general morphisms of quasi-projective varieties. We have discovered that the asymptotic ($p \to \infty$) behavior of the longest orbit of a fixed pseudo-Anosov g acting on $\mathbb{X}_{-2}(\mathbb{F}_p)$ is dictated by a dichotomy that we describe both in combinatorial terms and in algebraic terms related to Gauss's ambiguous binary quadratic forms, following Sarnak [23]. This dichotomy is illustrated with numerics, based on which we formulate a precise conjecture in Conjecture 1.10.

1. Introduction

For $\kappa \in \mathbf{Z}$, let $\mathbb{X}_\kappa$ denote the affine surface

$$\mathbb{X}_\kappa : x^2 + y^2 + z^2 = xyz + 2 + \kappa. \tag{1.1}$$

When $\kappa = -2$, $\mathbb{X}_{-2}$ is Markoff's surface. A theorem of Markoff [17] relates the integer points on $\mathbb{X}_{-2}$ to the Diophantine properties of $\mathbf{Q}$; in particular to the Markoff spectrum. In a different vein, the real and complex points of $\mathbb{X}_\kappa$ are related to moduli spaces of $\mathrm{SL}_2(\mathbf{C})$-local systems on a torus with one puncture [12]. Due to this connection, letting $\mathbf{F}_2$ denote the free group on 2 generators, the group $\mathrm{Out}(\mathbf{F}_2) \cong \mathrm{GL}_2(\mathbf{Z})$ acts by automorphisms of $\mathbb{X}_\kappa$, viewed as a scheme of finite type over $\mathbf{Z}$. We give a detailed description of this group action in Section 2.2 below.

The group $\mathrm{Out}(\mathbf{F}_2)$ is the mapping class group of the torus with one puncture, and the free group $\mathbf{F}_2$ is the fundamental group of this surface. As such, $\mathrm{Out}(\mathbf{F}_2)$ is subject to Thurston's classification of mapping class group elements [27] into periodic, reducible, or *pseudo-Anosov* (p-A.) elements. From the point of view of

M. Magee was supported in part by N.S.F. award DMS-1701357. All authors were supported in part by Sam Payne's N.S.F. CAREER award DMS–1149054.

$\mathrm{GL}_2(\mathbf{Z})$, an element is p-A. if it is *hyperbolic*, that is, has two distinct real eigenvalues. The current paper aims to investigate how p-A. elements of $\mathrm{Out}(\mathbf{F}_2)$ act on $\mathbb{X}_\kappa(\mathbb{F}_p)$ for prime p.

The study of p-A. elements of $\mathrm{Out}(\mathbf{F}_2)$ acting on $\mathbb{X}_\kappa(\mathbf{R})$ and $\mathbb{X}_\kappa(\mathbf{C})$ has been ongoing since the early 1980s, instigated by a paper of Kohmoto, Kadanoff and Tang [15] where the spectrum of a 1D lattice Schrödinger operator with a quasiperiodic potential was related to the dynamics of a particular p-A. automorphism (the *Fibonacci substitution*) on $\mathbb{X}_\kappa(\mathbf{R})$. In [4], Cantat resolved a conjecture of Kadanoff relating the topological entropy of a p-A. element acting on $\mathbb{X}_\kappa(\mathbf{R})$ to the largest eigenvalue of the corresponding matrix in $\mathrm{GL}_2(\mathbf{Z})$. See also Bowditch [3] for some related questions.

Here, we begin a parallel study for the action of p-A. elements on $\mathbb{X}_\kappa(\mathbb{F}_p)$. Any p-A. element Φ of $\mathrm{Out}(\mathbf{F}_2)$ gives for each prime p a permutation Φ_p of $\mathbb{X}_\kappa(\mathbb{F}_p)$. In this paper we propose that in the study of p-A. Φ acting on $\mathbb{X}_\kappa(\mathbb{F}_p)$, one should replace topological entropy by the asymptotic complexity of the family of permutations $\{\Phi_p\}$. In particular, we ask the following question.

Problem 1.1. *For fixed p-A. Φ, what is the asymptotic behavior of*

$$\frac{\log(\text{ longest cycle of } \Phi_p \text{ on } \mathbb{X}_\kappa(\mathbb{F}_p)\ )}{\log p}$$

as $p \to \infty$?

Empirically, the answer to this question is quite surprising (see Conjecture 1.10 below). We also obtain a theoretical result towards this question in Theorem 1.5 below. We will be primarily interested in the case of $\kappa = -2$, although we prove some of our results for general κ.

Before tackling Problem 1.1, a preliminary question intervenes. It could a priori be the case that the permutation group generated by $\mathrm{Out}(\mathbf{F}_2)$ on $\mathbb{X}_\kappa(\mathbb{F}_p)$ is highly restricted and this would of course affect how a single element can behave.

Let $\mathbb{X}^*_{-2}(\mathbb{F}_p) = \mathbb{X}_{-2}(\mathbb{F}_p) - (0,0,0)$. Bourgain, Gamburd, and Sarnak prove in [2, Theorem 2] that for all primes outside a very small exceptional set, the action of $\mathrm{Out}(\mathbf{F}_2)$ on $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ is transitive, which was a conjecture of McCullough and Wanderley from [18]. Sarnak has raised more generally the question of what permutation group is generated by the action of $\mathrm{Out}(\mathbf{F}_2)$ on $\mathbb{X}^*_{-2}(\mathbb{F}_p)$.

It follows from work of Horowitz [14] (see also Goldman [12]) that

$$\mathrm{Aut}(\mathbb{X}_\kappa) \cong \mathrm{PGL}_2(\mathbf{Z}) \ltimes N$$

where the $\mathrm{PGL}_2(\mathbf{Z})$ factor is induced by $\mathrm{Out}(\mathbf{F}_2)$ and N is the Klein four-group generated by even sign changes

$$n_1 : (x, y, z) \longmapsto (x, -y, -z)$$

(similarly n_2, n_3). For p odd, each N-orbit on $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ contains four distinct points (see Lemma 5.1 below). Thus $\mathrm{Out}(\mathbf{F}_2)$ cannot act 2-transitively on $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ for any prime, since it must permute orbits of N. In light of this observation, we should examine instead the action of $\mathrm{Out}(\mathbf{F}_2)$ on the set of N-orbits in $\mathbb{X}^*_{-2}(\mathbb{F}_p)$, which we denote by $\mathbb{Y}_{-2}(\mathbb{F}_p)$.

Let $H(p)$ denote the permutation group generated by $\mathrm{Out}(\mathbf{F}_2)$ acting on $\mathbb{Y}_{-2}(\mathbb{F}_p)$. We write A_n for the alternating group on n letters and S_n for the symmetric group. We prove the following.

Theorem 1.2. *Let $n = |\mathbb{Y}_{-2}(\mathbb{F}_p)|$ and let $p > 3$. Then, $H(p) \leq A_n$ if and only if $p \equiv 3 \pmod{16}$.*

We prove Theorem 1.2 in Section 5. Theorem 1.2, alongside computations of $H(p)$ for $p \leq 47$, lead us to conjecture the following:

Conjecture 1.3. *Let $H(p)$ denote the permutation group induced by the action of $\mathrm{Out}(\mathbf{F}_2)$ on $\mathbb{Y}_{-2}(\mathbb{F}_p)$, and let $n = |\mathbb{Y}_{-2}(\mathbb{F}_p)|$. Then when $p > 3$*

- $H(p) \cong S_n$ if $p \not\equiv 3 \pmod{16}$,
- $H(p) \cong A_n$ if $p \equiv 3 \pmod{16}$.

Meiri and Puder have proved in [19] that $H(p)$ contains A_n whenever $p \equiv 1 \pmod 4$ and p is outside the Bourgain-Gamburd-Sarnak exceptional set, and also for a density 1 set of primes without any congruence condition. In these cases, Theorem 1.2 describes exactly what $H(p)$ is. This shows there is no obstruction at the level of the group $H(p)$ to a p-A. element behaving 'generically' on $\mathbb{Y}_{-2}(\mathbb{F}_p)$.

We now describe our theoretical result towards Problem 1.1. In [26] Silverman studied a more general version of this problem and obtained as a consequence the following result.

Theorem 1.4 (Silverman [26, Theorem 3(a)])**.** *Let K be a number field with ring of integers $\mathcal{O}_K$ and V a quasi-projective variety defined over K. Let $\varphi : V \to V$ be a morphism defined over K, such that φ has an infinite orbit on $V(K)$. If $\mathfrak{p} \in \mathrm{Spec}(\mathcal{O}_K)$ is such that V and φ have good reduction at $\mathfrak{p}$, then write $\varphi_\mathfrak{p}$ and $V(\mathbb{F}_\mathfrak{p})$ for these reductions and $N(\mathfrak{p})$ for the norm of this prime. For any $\epsilon > 0$, the set of $\mathfrak{p} \in \mathrm{Spec}(\mathcal{O}_K)$ such that there is good reduction of φ and V at $\mathfrak{p}$, and an orbit of $\varphi_\mathfrak{p}$ on $V(\mathbb{F}_\mathfrak{p})$ of length $\geq (\log N(\mathfrak{p}))^{1-\epsilon}$ has analytic density 1.*

This result applies directly to Problem 1.1. Silverman's result actually provides many orbits of length $\geq (\log N(\mathfrak{p}))^{1-\epsilon}$.

What we can achieve in the current context is the removal of the ϵ from Theorem 1.4, and get a statement for all primes p instead of just analytic density 1. Furthermore, our bounds are independent of κ.

Theorem 1.5. *Given a pseudo-Anosov $g \in \mathrm{Out}(\mathbf{F}_2)$, let λ denote the eigenvalue of largest modulus of the corresponding matrix in $\mathrm{GL}_2(\mathbf{Z})$. For any $\kappa \in \mathbf{Z}$, as $p \to \infty$, g has an orbit of length at least*

$$\frac{\log p}{\log |\lambda|} + O_g(1)$$

on $\mathbb{X}_\kappa(\mathbb{F}_p)$. The implied constant depends on g, but not on κ.

Now we describe our numerical results which show what the answer to Problem 1.1 should be, at least for $\mathbb{Y}_{-2}(\mathbb{F}_p)$. First we give two natural guesses, that turn out to both be wrong.

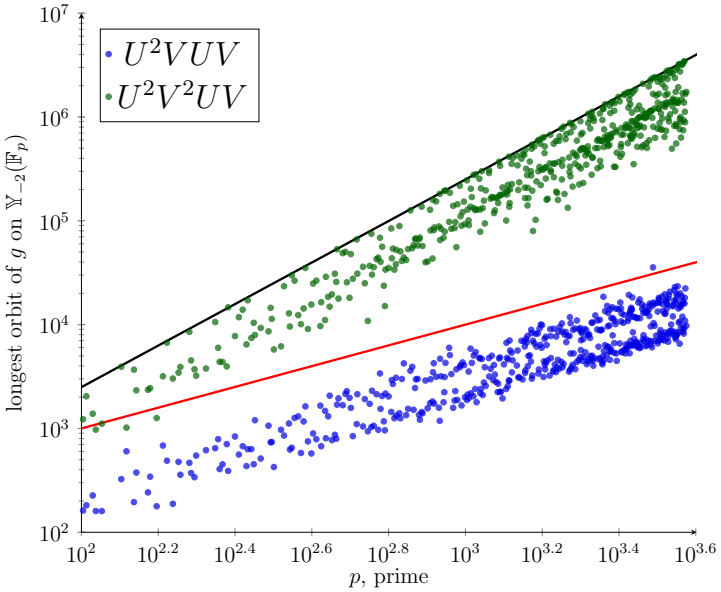

FIGURE 1.1. This shows the longest orbits of two p-A. elements U^2VUV (blue) and U^2V^2UV (green) on $\mathbb{Y}_{-2}(\mathbb{F}_p)$. The black line is $y = p^2/4$, which is asymptotic to $|\mathbb{Y}_{-2}(\mathbb{F}_p)|$. The red line is $y = 10p$. The plots are in log vs log scale axes with p on the horizontal axis and longest orbit on the vertical.

Guess 1: a p-A. g acts as a random map on $\mathbb{Y}_{-2}(\mathbb{F}_p)$. A random map from a set of size N to itself has with high probability its longest orbit of size $\asymp \sqrt{N}$. Since $|\mathbb{Y}_{-2}(\mathbb{F}_p)| \asymp p^2$ this predicts the longest orbit of g acting on $\mathbb{Y}_{-2}(\mathbb{F}_p)$ will have size $\asymp p$. This fact comes from a collision heuristic based on the 'Birthday paradox'. However, this heuristic is not convincing, since g is invertible, so should really be viewed as a random permutation (for some notion of random, see next guess). On the other hand, although this guess doesn't give the right answer in general (see below), there are p-A. g for which this guess does point to the right asymptotic behavior.

Guess 2: a p-A. g acts as a random permutation on $\mathbb{Y}_{-2}(\mathbb{F}_p)$. Perhaps we should model the action of g on $\mathbb{Y}_{-2}(\mathbb{F}_p)$ by a permutation chosen uniformly at random from A_n or S_n, where $n = |\mathbb{Y}_{-2}(\mathbb{F}_p)|$, according to Conjecture 1.3. To simplify things, let us just consider S_n, the case of A_n being similar. Then it is known that a permutation drawn uniformly at random from S_n has a cycle of length at least $n/2$ in its cycle decomposition with positive probability. This fact is closely related to the well-known '100 Prisoners Problem' posed in [10]. So this would predict for fixed p-A. g that as p varies we should often (in fact being more careful with the statistics, with high probability) see an orbit of length $\asymp p^2$ of g on $\mathbb{Y}_{-2}(\mathbb{F}_p)$. This guess also turns out not to be correct in general.

To describe our numerics, we introduce special elements of $\mathrm{PGL}_2(\mathbf{Z})$. We first note, if g has determinant -1, then the qualitative behavior of the longest orbit of g will be governed by that of g^2, which has determinant 1. So it is sufficient (at least for the phenomena we show) to consider only elements of $\mathrm{PSL}_2(\mathbf{Z})$. Let

$$U = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Lemma 1.6. *Every hyperbolic element of* $\mathrm{PSL}_2(\mathbf{Z})$ *is conjugate to an element*

$$U^{n_1}V^{m_1}\dots U^{n_k}V^{m_k}$$

with $k > 0$ *and all* $n_i, m_i > 0$. *We call this a reduced* UV*-word.*

This is well-known, but for completeness we prove Lemma 1.6 in Section 4 below. Since the orbit lengths of g on $\mathbb{X}_\kappa(\mathbb{F}_p)$ are the same after conjugation, we may simply consider reduced UV-words in what follows. Our conjectural answer to Problem 1.1 is based on a dichotomy for hyperbolic $g \in \mathrm{PSL}_2(\mathbf{Z})$.

Definition 1.7. A reduced UV-word $U^{n_1}V^{m_1}\dots U^{n_k}V^{m_k}$ is a *cyclic palindrome* if its reverse can be cyclically rotated to obtain the original word. For example:

$$U^2VUV \xrightarrow{\text{reverse}} VUVU^2 \xrightarrow{\text{rotate}} UVU^2V \xrightarrow{\text{rotate}} VU^2VU \xrightarrow{\text{rotate}} U^2VUV.$$

Then U^2VUV is a cyclic palindrome, whereas U^2V^2UV is not. Following Sarnak [23] (who follows terminology of Gauss) we make the following definition.

Definition 1.8. Say $g \in \mathrm{PSL}_2(\mathbf{Z})$ is *ambiguous* if the conjugacy class of g in $\mathrm{PSL}_2(\mathbf{Z})$ is conjugated to the conjugacy class of g^{-1} in $\mathrm{PSL}_2(\mathbf{Z})$ by an element of $\mathrm{PGL}_2(\mathbf{Z})$ of determinant -1.

Our two definitions actually coincide.

Proposition 1.9. *Let hyperbolic* $g \in \mathrm{PSL}_2(\mathbf{Z})$ *be given by a reduced* UV*-word. Then the* UV*-word is a cyclic palindrome if and only if* g *is ambiguous.*

We prove Proposition 1.9 in Section 4. In Figure 1.1 we show the longest orbits of U^2VUV and U^2V^2UV on $\mathbb{Y}_{-2}(\mathbb{F}_p)$. They evidently have strikingly different behaviors. Note that

$$U^2VUV = \begin{pmatrix} 2 & 3 \\ 5 & 8 \end{pmatrix}, \quad U^2V^2UV = \begin{pmatrix} 3 & 5 \\ 7 & 12 \end{pmatrix},$$

so they are both hyperbolic. However, Figure 1.1 shows that the longest orbit of U^2VUV is on the order of p and that of U^2V^2UV is on the order of p^2. Based on further evidence (see Table 1 and Figure 1.2), we are led to conjecture that the crucial difference between these words is that U^2VUV is a cyclic palindrome/ambiguous. Write $L(g;p)$ for the longest orbit of g on $\mathbb{Y}_{-2}(\mathbb{F}_p)$. We make the following conjecture:

Conjecture 1.10. *Let $g \in \mathrm{PSL}_2(\mathbf{Z})$ be hyperbolic. If g is ambiguous then*

(1) *There are constants $C_1 = C_1(g) > 0$ and $C_2 = C_2(g) > C_1$ such that $C_1 p \leq L(g;p) \leq C_2 p$ for all primes p.*
(2) *The discrete probability measures*

$$\frac{1}{\#\{primes\ p \leq X\}} \left(\sum_{p \leq X} \delta_{\frac{L(g;p)}{p}} \right)$$

converge as $X \to \infty$ to a compactly supported Borel probability measure on $\mathbf{R}$.

*If g is **not** ambiguous then*

(1) *There is a constant $c = c(g)$ such that $L(g;p) \geq cp^2$ for all primes p.*
(2) *The discrete probability measures*

$$\frac{1}{\#\{primes\ p \leq X\}} \left(\sum_{p \leq X} \delta_{\frac{L(g;p)}{p^2}} \right)$$

converge as $X \to \infty$ to a compactly supported Borel probability measure on $\mathbf{R}$.

As a particular consequence, we conjecture that the answer to Problem 1.1 for $\kappa = -2$ is

$$\lim_{p \to \infty} \frac{\log(L(g;p))}{\log p} = \begin{cases} 1 & \text{if } g \text{ is ambiguous.} \\ 2 & \text{if } g \text{ is not ambiguous.} \end{cases}$$

The issue of whether elements of $\mathrm{SL}_2(\mathbf{Z})$ are conjugate to their inverses shows up in several different areas of mathematics including connect sum problems for manifolds [7], the dynamics of kicked toral automorphisms [22], and the classification of foliations of torus bundles over the circle [11]. This issue is explored in depth in the article of Sarnak [23] where it is related to the theory of binary quadratic forms. A conjugacy class in $\mathrm{PSL}_2(\mathbf{Z})$ is called *primitive* if a representative is not a power of another element. To each conjugacy class $[g]$ in $\mathrm{PSL}_2(\mathbf{Z})$ one can attach a number $t([g]) = |\mathrm{trace}(g)|$. Let Π denote the collection of primitive hyperbolic conjugacy classes in $\mathrm{PSL}_2(\mathbf{Z})$. It is a result of Hejhal [13], after Selberg [25], that one has the asymptotic formula

$$\sum_{p \in \Pi,\, t(p) \leq X} 1 \approx \frac{X^2}{2 \log X}.$$

On the other hand, Sarnak shows in [23] that if we write Π_A for the collection of primitive hyperbolic *ambiguous* conjugacy classes in $\mathrm{PSL}_2(\mathbf{Z})$, then

$$\sum_{p \in \Pi_A,\, t(p) \leq X} 1 \approx \frac{97}{8\pi^2} X (\log X)^2.$$

So the ambiguous classes are rare, with those having $t(p) \leq X$ taking up about a square root of the number of all primitive hyperbolic classes with $t(p) \leq X$.

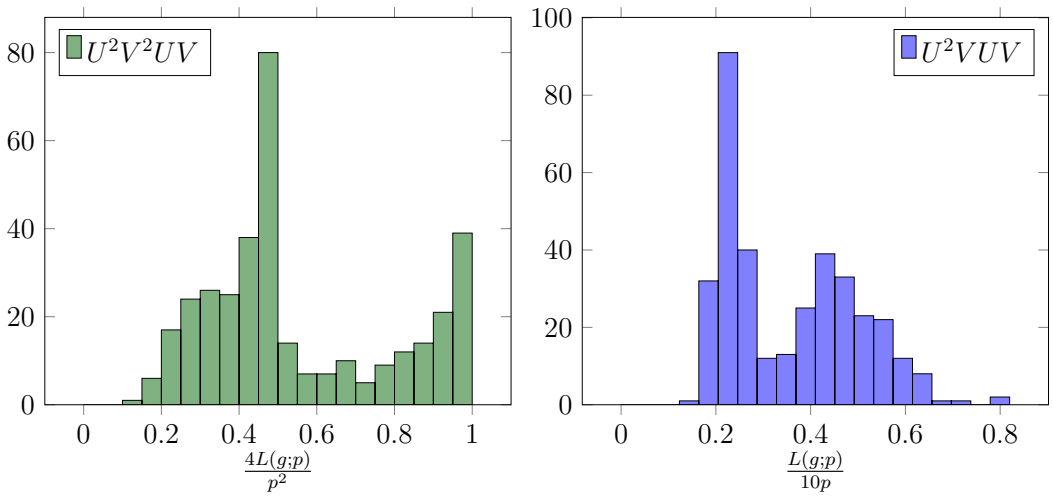

FIGURE 1.2. Histograms showing the distributions that feature in Conjecture 1.10. Here p ranges between 1009 and 3761. For U^2V^2UV the histogram shows the distribution of $L(g;p)(p^2/4)^{-1}$. Note $|\mathbb{Y}_{-2}(\mathbb{F}_p)|$ is asymptotic to $p^2/4$. For U^2VUV the distribution is of $L(g;p)(10p)^{-1}$. The 10 is not significant and has just been chosen to scale the data. One outlier ($p = 3079$, $L(g;p) = 35585$) has been removed from the U^2VUV chart.

Acknowledgments. We thank Matthew de Courcy-Ireland, Alexei Entin, Alex Gamburd, José Gonzalez, Sam Payne, Doron Puder, Dhruv Ranganathan, Peter Sarnak, and Joseph Silverman for enlightening discussions about this work. We also thank the anonymous referee whose suggestions have improved this paper.

The first version of this paper was written during the Summer Undergraduate Research at Yale program, funded in part by Sam Payne's N.S.F. CAREER award DMS–1149054. M. Magee was supported in part by N.S.F. award DMS-1701357.

2. Background

2.1. **Background on the free group.** Here we give necessary background about the free group $\mathbf{F}_2$ and its automorphisms. We fix generators X and Y of $\mathbf{F}_2$ throughout this paper. We always assume words in $\mathbf{F}_2$ are reduced, meaning positive powers of X do not appear beside negative powers, and similarly for Y. Following [21] we make the following definition.

Definition 2.1. A word $w \in \mathbf{F}_2$ is *monotone* if for each letter X or Y, all the exponents of this letter in w have the same sign. For example, XY^{-1} is monotone, but YXY^{-1} is not.

We need the following proposition that appears in Parzanchevski and Puder [21, Prop. 3.5].

g **not** ambiguous		g ambiguous	
g	$L(g;727)$	g	$L(g;727)$
$V^1U^1V^3U^1V^2U^2$	87928	$V^1U^1V^1U^1V^1U^2$	3193
$V^3U^2V^1U^2V^2U^2$	77996	$V^1U^1V^3U^1V^1U^2$	2018
$V^1U^1V^2U^1V^2U^3$	75289	$V^2U^1V^2U^3V^2U^1$	2780
$V^2U^1V^1U^2V^1U^2$	95183	$V^4U^1V^1U^2V^1U^1$	3748
$V^2U^1V^1U^1V^3U^1$	42238	$V^1U^2V^1U^2V^3U^2$	2780
$V^2U^1V^1U^2V^2U^3$	62702	$V^1U^1V^1U^1V^1U^4$	2894
$V^1U^1V^1U^3V^2U^1$	51981	$V^1U^1V^1U^2V^1U^2$	4591
$V^1U^1V^3U^4V^1U^1$	75716	$V^1U^3V^1U^1V^1U^1$	3285
$V^1U^4V^2U^1V^1U^1$	79495	$V^1U^2V^1U^2V^1U^2$	3331
$V^1U^3V^2U^2V^3U^1$	86897	$V^2U^2V^2U^1V^2U^2$	3350
$V^3U^1V^1U^2V^1U^3$	108710	$V^1U^4V^1U^1V^4U^1$	1756
$V^2U^3V^1U^1V^3U^1$	61549	$V^2U^1V^2U^4V^2U^1$	2022
$V^1U^1V^2U^4V^3U^1$	87870	$V^2U^1V^1U^1V^2U^4$	2937
$V^1U^1V^2U^1V^3U^2$	82633	$V^1U^2V^1U^1V^1U^1$	3193
$V^2U^4V^1U^1V^1U^1$	79495	$V^1U^2V^2U^2V^1U^1$	3680
$V^4U^1V^1U^1V^1U^4$	130737	$V^1U^2V^3U^2V^1U^2$	2780
$V^3U^4V^1U^1V^2U^1$	72046	$V^1U^2V^1U^1V^4U^1$	3748

TABLE 1. This table gives evidence for Conjecture 1.10. The data is for $p = 727$. We have $|\mathbb{Y}_{-2}(\mathbb{F}_{727})| = 131587$. Recall $L(g;727)$ is the longest orbit of g on $\mathbb{Y}_{-2}(\mathbb{F}_{727})$.

Proposition 2.2. *Any element g of $\mathrm{Out}(\mathbf{F}_2)$ has a representative in $\mathrm{Aut}(\mathbf{F}_2)$ (modulo conjugation) of the form*

$$\hat{g} : (X, Y) \mapsto (w_1, w_2) \tag{2.1}$$

where w_1 and w_2 are monotone words in $\mathbf{F}_2$.

In the setting of Proposition 2.2 we say that $\hat{g}$ is *monotone.* Suppose $\hat{g} \in \mathrm{Aut}(\mathbf{F}_2)$ as in (2.1) is monotone, with

$$w_i = X^{\alpha_1^i} Y^{\beta_1^i} X^{\alpha_2^i} Y^{\beta_2^i} \dots X^{\alpha_{t_i}^i} Y^{\beta_{t_i}^i}$$

for some $\alpha_j^i, \beta_j^i, t_i \in \mathbf{Z}$. We identify $\mathbf{Z}^2 \cong \mathbf{F}_2/[\mathbf{F}_2, \mathbf{F}_2]$ by the basis induced by X, Y. Then $\hat{g}$ acts on $\mathbf{Z}^2$ by the matrix

$$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \in \mathrm{GL}_2(\mathbf{Z})$$

where $a_i = \sum_j \alpha_j^i$ and $b_i = \sum_j \beta_j^i$. This matrix clearly only depends on $g \in \mathrm{Out}(\mathbf{F}_2)$. Thus we obtain a homomorphism

$$\mathrm{Out}(\mathbf{F}_2) \to \mathrm{GL}_2(\mathbf{Z});$$

in fact, by a result of Nielsen [20], this map is an isomorphism, so the values a_i and b_i are uniquely determined by $g \in \mathrm{Out}(\mathbf{F}_2)$ and vice versa. We pass freely between these representations of g in the rest of the paper.

2.2. **Automorphisms of Markoff's surface.** The group actions that we consider in this paper arise from *representation varieties*. Here we follow Goldman [12]. We write $\mathrm{Hom}(\mathbf{F}_2, \mathrm{SL}_2(\mathbf{C}))$ for the set of homomorphisms from $\mathbf{F}_2$ to $\mathrm{SL}_2(\mathbf{C})$. We view an element $\phi \in \mathrm{Hom}(\mathbf{F}_2, \mathrm{SL}_2(\mathbf{C}))$ as an element of $\mathbf{C}^8$ via the coordinates of $\phi(X), \phi(Y)$ and hence view $\mathrm{Hom}(\mathbf{F}_2, \mathrm{SL}_2(\mathbf{C}))$ as an affine variety. There is an algebraic action of $\mathrm{SL}_2(\mathbf{C})$ on $\mathrm{Hom}(\mathbf{F}_2, \mathrm{SL}_2(\mathbf{C}))$ by conjugation, and a commuting action of $\mathrm{Aut}(\mathbf{F}_2)$.

Let $\mathcal{O} = \mathcal{O}_{\mathrm{Hom}(\mathbf{F}_2, \mathrm{SL}_2(\mathbf{C}))}$ denote the coordinate ring of $\mathrm{Hom}(\mathbf{F}_2, \mathrm{SL}_2(\mathbf{C}))$. We consider the ring of invariant functions $\mathcal{O}^{\mathrm{SL}_2(\mathbf{C})}$. By results of Fricke [8] and Fricke-Klein [9], $\mathcal{O}^{\mathrm{SL}_2(\mathbf{C})} \cong \mathbf{C}[x, y, z]$ with $x = \mathrm{tr}\phi(X)$, $y = \mathrm{tr}\phi(Y)$, $z = \mathrm{tr}\phi(XY)$ (considered as functions of $\phi \in \mathrm{Hom}(\mathbf{F}_2, \mathrm{SL}_2(\mathbf{C}))$). The action of $\mathrm{Aut}(\mathbf{F}_2)$ on $\mathcal{O}$ descends to an action of $\mathrm{Out}(\mathbf{F}_2)$ on $\mathcal{O}^{\mathrm{SL}_2(\mathbf{C})}$ by polynomial maps.

Now one has the following unexpected fact due to Nielsen [20]: the action of $\mathrm{Out}(\mathbf{F}_2)$ on conjugacy classes in $\mathbf{F}_2$ leaves invariant the pair of conjugacy classes given by the commutator $[X, Y]$ and its inverse. This, together with the fact that if $A \in \mathrm{SL}_2(\mathbf{C})$ then $\mathrm{tr}(A) = \mathrm{tr}(A^{-1})$, implies that $\mathrm{Out}(\mathbf{F}_2)$ preserves $\mathrm{tr}\phi([X, Y]) \in \mathcal{O}^{\mathrm{SL}_2(\mathbf{C})}$. Moreover, an identity of Fricke and Klein states

$$\mathrm{tr}\phi([X, Y]) + 2 = (\mathrm{tr}\phi(X))^2 + (\mathrm{tr}\phi(Y))^2 + (\mathrm{tr}\phi(XY))^2 - \mathrm{tr}\phi(X)\mathrm{tr}\phi(Y)\mathrm{tr}\phi(XY).$$

This explains that $\mathrm{GL}_2(\mathbf{Z})$ acts by polynomial automorphisms of both $\mathbf{C}^3$ and $\mathbf{C}[x, y, z]$ and preserves the polynomial

$$x^2 + y^2 + z^2 - xyz. \tag{2.2}$$

The center of $\mathrm{Out}(\mathbf{F}_2) = \mathrm{GL}_2(\mathbf{Z})$ is generated by the class of the automorphism $(X, Y) \mapsto (X^{-1}, Y^{-1})$. This corresponds to the matrix $-\mathrm{Id} \in \mathrm{GL}_2(\mathbf{Z})$. This element acts trivially on $\mathbf{C}^3$ since for $\phi : \mathbf{F}_2 \to \mathrm{SL}_2(\mathbf{C})$

$$(\mathrm{tr}\phi(X^{-1}), \mathrm{tr}\phi(Y^{-1}), \mathrm{tr}\phi(X^{-1}Y^{-1})) = (\mathrm{tr}\phi(X), \mathrm{tr}\phi(Y), \mathrm{tr}\phi(XY)).$$

Hence the action of $\mathrm{GL}_2(\mathbf{Z})$ on $\mathbf{C}^3$ factors through one of $\mathrm{PGL}_2(\mathbf{Z})$. Moreover, $\mathrm{PGL}_2(\mathbf{Z})$ is generated by standard unipotent matrices U and V defined in the Introduction together with

$$\eta := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The matrices U and V have representative automorphisms $\tilde{U}(X, Y) = (XY, Y)$, $\tilde{V}(X, Y) = (X, XY)$ in $\mathrm{Aut}(\mathbf{F}_2)$ from which one can work out the corresponding

polynomial maps of $\mathbf{C}^3$. Indeed, from the Cayley-Hamilton Theorem, for any $A, B \in \mathrm{SL}_2(\mathbf{C})$ we have

$$\mathrm{tr}(AB) = \mathrm{tr}(A)\mathrm{tr}(B) - \mathrm{tr}(AB^{-1}). \tag{2.3}$$

Hence we have for $x_0 = \mathrm{tr}\phi(X), y_0 = \mathrm{tr}\phi(Y), z_0 = \mathrm{tr}\phi(XY)$,

$$U(x_0, y_0, z_0) = (\mathrm{tr}\phi(\tilde{U}^{-1}(X)), \mathrm{tr}\phi(\tilde{U}^{-1}(Y)), \mathrm{tr}\phi(\tilde{U}^{-1}(XY)))$$

and since $\tilde{U}^{-1}(X, Y) = (XY^{-1}, Y)$ we obtain

$$U(x_0, y_0, z_0) = (\mathrm{tr}\phi(XY^{-1})), \mathrm{tr}\phi(Y), \mathrm{tr}\phi(X)) = (x_0y_0 - z_0, y_0, x_0). \tag{2.4}$$

Similar calculations give that

$$U^{-1}(x_0, y_0, z_0) = (z_0, y_0, y_0z_0 - x_0), \tag{2.5}$$

$$V(x_0, y_0, z_0) = (x_0, x_0y_0 - z_0, y_0), \quad V^{-1}(x_0, y_0, z_0) = (x_0, z_0, x_0z_0 - y_0), \tag{2.6}$$

$$\eta(x_0, y_0, z_0) = \eta^{-1}(x_0, y_0, z_0) = (x_0, y_0, x_0y_0 - z_0). \tag{2.7}$$

In particular, the action of $\mathrm{GL}_2(\mathbf{Z})$ on $\mathbf{C}^3$ preserves $\mathbf{Z}^3$.

We now consider the affine scheme over $\mathbf{Z}[\kappa]$

$$\mathbb{X} := \mathrm{Spec}(R)$$

where

$$R := \mathbf{Z}[\kappa, x, y, z]/I, \quad I := (x^2 + y^2 + z^2 - xyz - 2 - \kappa).$$

For particular choice of $\kappa \in \mathbf{Z}$ we obtain a scheme over $\mathbf{Z}$ that we denote by

$$\mathbb{X}_\kappa := \mathrm{Spec}(R_\kappa),$$

where

$$R_\kappa := \mathbf{Z}[x, y, z]/I_\kappa, \quad I_\kappa := (x^2 + y^2 + z^2 - xyz - 2 - \kappa).$$

In the case of $\kappa = -2$ one obtains Markoff's surface $\mathbb{X}_{-2}$. By the previous discussion, $\mathrm{Out}(\mathbf{F}_2) \cong \mathrm{GL}_2(\mathbf{Z})$ acts on $\mathbb{X}$ by automorphisms of schemes over $\mathbf{Z}[\kappa]$ and for each κ, $\mathrm{GL}_2(\mathbf{Z})$ acts on $\mathbb{X}_\kappa$ by automorphisms.

Given $g \in \mathrm{GL}_2(\mathbf{Z})$ we write $g_*(x), g_*(y), g_*(z) \in \mathbf{Z}[x, y, z]$ for the pushforwards of the generators of $\mathbf{Z}[x, y, z]$ under g. If $\phi \in \mathrm{Hom}(\mathbf{F}_2, \mathrm{SL}_2(\mathbf{C}))$ then

$$\phi \mapsto g(\phi)$$

in the coordinates $x_0 = \mathrm{tr}\phi(X), y_0 = \mathrm{tr}\phi(Y), z_0 = \mathrm{tr}\phi(XY)$ corresponds to

$$(x_0, y_0, z_0) \mapsto g(x_0, y_0, z_0) = (g_*^{-1}(x), g_*^{-1}(y), g_*^{-1}(z))|_{(x_0,y_0,z_0)}.$$

The inverses here are important to note; they cater to the distinction between the group action on *coordinate functions* and the group action on *points* of the scheme.

Aside from those already mentioned, there are other polynomial automorphisms of $\mathbf{Z}^3$ preserving the polynomial (2.2) and hence acting on $\mathbb{X}$ and $\mathbb{X}_\kappa$. These will be used in the sequel so we introduce them now. Let m_i denote the *Markoff moves* defined for $i = 1, 2, 3$ by

$$m_1(x, y, z) = (yz - x, y, z),\ m_2(x, y, z) = (x, xz - y, z),\ m_3(x, y, z) = (x, y, xy - z)$$

These moves correspond to fixing two coordinates in (2.2) and flipping the root of the quadratic equation in the remaining coordinate, hence they are sometimes called *Vieta involutions*. One also has an action of S_3 on $\mathbf{Z}^3$ by permuting coordinates, that clearly preserves (2.2). The Markoff moves and the action by S_3 are contained in, and generate, the image of $\mathrm{PGL}_2(\mathbf{Z})$ in $\mathrm{Aut}(\mathbb{X}_\kappa)$. As described in the Introduction, there is also a finite abelian group N of even sign changes preserving (2.2). These are *not* induced by the $\mathrm{GL}_2(\mathbf{Z})$ action, and a result of Horowitz [14] says that N and $\mathrm{GL}_2(\mathbf{Z})$ generate all complex polynomial automorphisms of $\mathbb{X}_\kappa(\mathbf{C})$.

2.3. **The Cayley Cubic.** When $\kappa = 2$, $\mathbb{X}_2$ is *Cayley's cubic surface* [6]. In fact $\mathbb{X}_2$ is closely related to the split torus $\mathbb{G}_m^2$; we heavily exploit this fact in the sequel. To see this, let $\tilde{\mathbb{X}}_2 := \mathrm{Spec}(\tilde{R}_2)$ where

$$\tilde{R}_2 := \mathbf{Z}[x,y,z,\delta,\eta]/J_2, \quad J_2 := (x^2+y^2+z^2-xyz-4, \delta^2 - x\delta + 1, \eta^2 - y\eta + 1).$$

The mapping

$$\begin{aligned} \tilde{R}_2 &\to \mathcal{O}_{\mathbb{G}_m^2} := \mathbf{Z}[\delta,\delta^*,\eta,\eta^*]/(\delta\delta^*-1, \eta\eta^*-1) \\ x &\mapsto \delta + \delta^* \qquad (2.8) \\ y &\mapsto \eta + \eta^* \qquad (2.9) \\ z &\mapsto \delta\eta + \delta^*\eta^* \qquad (2.10) \end{aligned}$$

and $\delta, \eta \mapsto \delta, \eta$ induces an isomorphism $\tilde{\mathbb{X}}_2 \cong \mathbb{G}_m^2$. The inclusion of $R_2 \to \tilde{R}_2$ induces a map

$$\mathbb{G}_m^2 \cong \tilde{\mathbb{X}}_2 \to \mathbb{X}_2.$$

Remark 2.3. *The existence of the mapping $\mathbb{G}_m^2 \to \mathbb{X}_2$ described above arises from the fact that $\mathbb{X}_2(\mathbf{C})$ can be identified with the collection of $\mathrm{SL}_2(\mathbf{C})$-conjugacy classes of semisimple reducible representations in $\mathrm{Hom}(\mathbf{F}_2, \mathrm{SL}_2(\mathbf{C}))$, or in simpler terms, conjugacy classes of pairs of diagonal matrices in $\mathrm{SL}_2(\mathbf{C})$. Although we do not use this fact, it is useful to keep in mind.*

There is an action of $\mathrm{GL}_2(\mathbf{Z})$ on $\mathcal{O}_{\mathbb{G}_m^2}$ by

$$\begin{aligned} g_*(\delta) = \delta^a\eta^c, \quad g_*(\eta) = \delta^b\eta^d, \\ g_*(\delta^*) = (\delta^*)^a(\eta^*)^c, \quad g_*(\eta^*) = (\delta^*)^b(\eta^*)^d, \qquad (2.11) \end{aligned}$$

for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbf{Z})$. We interpret $\delta^{-n} = (\delta^*)^n$ for $n \in \mathbf{Z}$ and similarly $\eta^{-n} = (\eta^*)^n$.

Let ι be the map $\iota : R_2 \to \mathcal{O}_{\mathbb{G}_m^2}$ defined by the inclusion $R_2 \to \tilde{R}_2$ followed by the map $\tilde{R}_2 \to \mathcal{O}_{\mathbb{G}_m^2}$ given by (2.8), (2.9), (2.10). This induces a map

$$\iota^* : \mathbb{G}_m^2 \to \mathbb{X}_2.$$

Lemma 2.4. *The map $\iota^* : \mathbb{G}_m^2 \to \mathbb{X}_2$ is $\mathrm{GL}_2(\mathbf{Z})$-equivariant.*

Proof. Recall U, V from our Introduction and η from Section 2.2. The lemma can be checked by noting that $\mathrm{GL}_2(\mathbf{Z})$ is generated by U, V and η, and these act on R_2 by

$$\begin{aligned}(U_*(x), U_*(y), U_*(z)) &= (z, y, yz - x),\\ (V_*(x), V_*(y), V_*(z)) &= (x, z, xz - y),\\ (\eta_*(x), \eta_*(y), \eta_*(z)) &= (x, y, xy - z).\end{aligned}$$

(cf. (2.4)–(2.7)). Then taking V as an example, $V = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, and therefore using (2.8), (2.9), (2.10) and (2.11) gives

$$\begin{aligned}(V_* \circ \iota(x), V_* \circ \iota(y), V_* \circ \iota(z)) &= (V_*(\delta + \delta^*), V_*(\eta + \eta^*), V_*(\delta\eta + \delta^*\eta^*))\\ &= (\delta + \delta^*, \delta\eta + \delta^*\eta^*, \delta^2\eta + (\delta^*)^2\eta^*)\\ &= (\iota(x), \iota(z), \iota(xz - y)) = (\iota V_*(x), \iota V_*(y), \iota V_*(z)).\end{aligned}$$

The calculations for U and η are similar. □

3. Lower bound on the longest orbit

3.1. **Proof of Theorem 1.5.** Our proof of Theorem 1.5 relies on proving that reasonably small powers of g have few fixed points. It is convenient for our analysis to introduce the following definition.

Definition 3.1 (Good matrices)**.** Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbf{Z})$. We say g is *good* if $a, b, c, d \geq 2$.

We use the notation $O_x(x^n)$ for the class of polynomials containing terms with x-degree $\leq n$, that is, with no monomial summand containing a power of x greater than n.

The main technical ingredient in the proof of Theorem 1.5 is the following result, whose proof will be deferred to Section 3.2.

Proposition 3.2. *For each coprime $a, c \in \mathbf{Z}$ with $a \geq 2$, $c \geq 2$ there are polynomials $\tilde{p}_{a,c}$, $\tilde{q}_{a,c} \in \mathbf{Z}[x, y]$ with the following properties. Assume $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbf{Z})$ and that g is good. Recall that $I = (x^2 + y^2 + z^2 - xyz - 2 - \kappa) \subset \mathbf{Z}[\kappa, x, y, z]$.*

(1) *We have*

$$g_*(x) = \tilde{p}_{a,c} + \tilde{q}_{a,c} z \mod I, \quad g_*(y) = \tilde{p}_{b,d} + \tilde{q}_{b,d} z \mod I.$$

(2) *Moreover,*

$$\tilde{D} := \det \begin{pmatrix} \tilde{p}_{a,c} - x & \tilde{q}_{a,c} \\ \tilde{p}_{b,d} - y & \tilde{q}_{b,d} \end{pmatrix} \in \mathbf{Z}[\kappa, x, y]$$

is given by

$$\tilde{D} = x^{a+b-1} D_0 + O_x(x^{a+b-2})$$

where $D_0 \in \mathbf{Z}[y]$ *with* $D_0 \neq 0$ *and monic, up to a sign, and* $\deg(D_0) = |d-c|-1 \geq 0$.

With Proposition 3.2 in hand, we can control fixed points:

Lemma 3.3. *Let* $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbf{Z})$ *with* $|a|, |b|, |c|, |d| \geq 2$, *then for any* $\kappa \in \mathbf{Z}$, g *has fewer than* $2p(||d|-|c||+|a|+|b|)$ *fixed points in* $\mathbb{X}_\kappa(\mathbb{F}_p)$.

Proof. First if $ab < 0$ then conjugating by $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ gives a new matrix which has $ab > 0$. Now if $a < 0$, multiplying by $-I$ gives a new matrix with $a, b, c, d \geq 2$, i.e. the resulting matrix is good. These operations do not change the conjugacy class of the matrix in $\mathrm{PGL}_2(\mathbf{Z})$, therefore the number of fixed points on $\mathbb{X}_\kappa(\mathbb{F}_p)$, and neither do they change the quantity $2p(||d|-|c||+|a|+|b|)$. So we may assume without loss of generality that g is good.

In this proof we distinguish a specific fixed value κ_0 from the generic parameter κ of $\mathbf{Z}[\kappa, x, y]$. Let $\tilde{p}_{a,c}, \tilde{q}_{a,c}$ be the polynomials from Proposition 3.2, and let $p_{a,c}^{\kappa_0}$, $q_{a,c}^{\kappa_0}$ be the images of $\tilde{p}_{a,c}$, $\tilde{q}_{a,c}$ under the evaluation map

$$\pi_{\kappa_0} : \mathbf{Z}[\kappa, x, y] \to \mathbf{Z}[x, y], \quad \kappa \mapsto \kappa_0.$$

If $(X, Y, Z) \in \mathbb{X}_{\kappa_0}(\mathbb{F}_p)$ is a fixed point of g, then from Proposition 3.2 we know $p_{a,c}^{\kappa_0}(X, Y) + q_{a,c}^{\kappa_0}(X, Y)Z = X$ and $p_{b,d}^{\kappa_0}(X, Y) + q_{b,d}^{\kappa_0}(X, Y)Z = Y$ so

$$\begin{pmatrix} p_{a,c}^{\kappa_0}(X,Y) - X & q_{a,c}^{\kappa_0}(X,Y) \\ p_{b,d}^{\kappa_0}(X,Y) - Y & q_{b,d}^{\kappa_0}(X,Y) \end{pmatrix} \begin{pmatrix} 1 \\ Z \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \bmod p. \tag{3.1}$$

In particular, the determinant $D^{\kappa_0} \in \mathbf{Z}[x, y]$ of this matrix must be zero when evaluated at $(X, Y) \in \mathbb{F}_p^2$. But, recalling Proposition 3.2 and its notation,

$$D^{\kappa_0} = \pi_{\kappa_0}(D) = \pi_{\kappa_0}\left(x^{a+b-1}D_0 + O_x(x^{a+b-2})\right) = x^{a+b-1}D_0 + O_x(x^{a+b-2})$$

by using that $D_0 \in \mathbf{Z}[y]$.

Proposition 3.2 tells us that for $Y \in \mathbb{F}_p$ with $D_0(Y) \neq 0$, the polynomial in $\mathbb{F}_p[x]$ obtained by evaluating D^{κ_0} at $y = Y$ has degree $a + b + 1$. Also from Proposition 3.2, D_0 is monic up to a sign with degree $|d-c|-1$, so there are at most

$$p.(||d|-|c||-1) + p(a+b+1) = p(||d|-|b||+|a|+|c|)$$

pairs $(X, Y) \in \mathbb{F}_p^2$ for which (3.1) can hold. On the other hand, since $X^2+Y^2+Z^2 = XYZ + 2 + \kappa_0$, given X, Y for which (3.1) holds, there are at most two possible Z with $(X, Y, Z) \in \mathbb{X}_{\kappa_0}(\mathbb{F}_p)$. □

Now that we have control over the number of fixed points of elements of $g \in \mathrm{GL}_2(\mathbf{Z})$ we can proceed to prove Theorem 1.5.

Proof of Theorem 1.5. Given hyperbolic g in $\mathrm{GL}_2(\mathbf{Z})$, we consider powers g^n of this element. Let λ be the eigenvalue of g of largest modulus. Diagonalizing g we have

$$g^n = \begin{pmatrix} Q_{11}(\lambda^n, \lambda^{-n}) & Q_{12}(\lambda^n, \lambda^{-n}) \\ Q_{21}(\lambda^n, \lambda^{-n}) & Q_{22}(\lambda^n, \lambda^{-n}) \end{pmatrix}$$

where the Q_{ij} are quadratic forms depending on g. It is possible to check that since g is hyperbolic, all the coefficients of g^n are unbounded as $n \to \infty$ in the sense that for all $M > 0$, there is $N(M)$ such that when $n > N(M)$, $|(g^n)_{ij}| > M$ for all $1 \leq i, j \leq 2$.

Indeed, if g is hyperbolic it cannot fix $[1;0]$ or $[0;1]$ in the action of $\mathrm{GL}_2(\mathbf{Z})$ on $P^1(\mathbf{R})$. So g has an attracting fixed point z_+ in $P^1(\mathbf{R})$ that is distinct from than $[1;0]$ and $[0;1]$. Of course the same is true for the transpose g^T. This means, projectively, g^n converges to a matrix with all entries nonzero. Since $\mathrm{GL}_2(\mathbf{Z})$ is discrete, at least one entry of g^n is unbounded, hence all the entries are.

Note this implies that for $n \geq n_0(g)$, g^n satisfies the hypothesis of Lemma 3.3. Noting that there is $C = C(g)$ such that all coefficients of g^n are $\leq C\lambda^n$, Lemma 3.3 gives that g^n has fewer than $8Cp|\lambda|^n$ fixed points on $\mathbb{X}_\kappa(\mathbb{F}_p)$ when $n \geq n_0(g)$.

We also need a bound on the number of fixed points of g^n when $n < n_0(g)$. In this case, we have that $g^{n\lceil n_0(g)/n \rceil}$ satisfies the hypothesis of Lemma 3.3, so it has fewer than Mp fixed points, where

$$M = 8 \max\{ \, |(g^{n\lceil n_0(g)/n \rceil})_{i,j}| \, : \, 1 \leq i, j \leq 2, \, , \, 1 \leq n < n_0(g) \, \}.$$

But any fixed point of g^n gives rise to a fixed point of $g^{n\lceil n_0(g)/n \rceil}$ so this means g^n has fewer than Mp fixed points.

For given N, this implies that the number of points in $\mathbb{X}_\kappa(\mathbb{F}_p)$ fixed by any g^n with $n \leq N$ is

$$\leq \sum_{n < n_0} Mp + \sum_{n_0 \leq n \leq N} 8Cp|\lambda|^n \leq n_0 Mp + C'p|\lambda|^N.$$

for $C' = C'(g) > 0$. We have $|\mathbb{X}_\kappa(\mathbb{F}_p)| \leq cp^2$ with c depending only on the complexity of $\mathbb{X}_\kappa$ viewed as a variety over $\mathbb{F}_p$, hence independent of κ. This follows from the Lang-Weil bound [16, Lemma 1], and also from direct consideration of (1.1). Therefore if $n_0 Mp + C'p|\lambda|^N < cp^2$ then there exists a point in $\mathbb{X}_\kappa(\mathbb{F}_p)$ not fixed by g^n for any $n \leq N$. Hence there is a cycle of g of length $\geq N$ where

$$N \approx \frac{\log\left(\frac{cp}{2C'} - \frac{n_0 M}{C'}\right)}{\log|\lambda|} = \frac{\log p}{\log|\lambda|} + O_g(1).$$

□

3.2. **Proof of Proposition 3.2.** In this section we prove the technical Proposition 3.2. The first step is to calculate the action of a given $g \in \mathrm{GL}_2(\mathbf{Z})$ on the Cayley cubic $\mathbb{X}_2$ and prove the analog of Proposition 3.2 when κ is evaluated at 2. We will do this by exploiting the embedding $\iota : R_2 \to \mathcal{O}_{\mathbb{G}_m^2}$.

Proposition 3.4. *For each coprime $a, c \in \mathbf{Z}$ with $a \geq 2$, $c \geq 2$ there are polynomials $p_{a,c}$, $q_{a,c} \in \mathbf{Z}[x,y]$ with the following properties. Assume $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbf{Z})$ and that g is good. Recall that $I_2 = (x^2+y^2+z^2-xyz-4) \subset \mathbf{Z}[x,y,z]$.*

(1) *We have*

$$g_*(x) = p_{a,c} + q_{a,c}z \mod I_2, \quad g_*(y) = p_{b,d} + q_{b,d}z \mod I_2$$

where $x, y \in R_2$ are the first two coordinate functions on $\mathbb{X}_2$. Here when we make statements that relate elements of $\mathbf{Z}[x,y]$ to elements of R_κ we always use the natural inclusion $\mathbf{Z}[x,y] \to R_\kappa$.

(2) *We have*

$$D := \det \begin{pmatrix} p_{a,c} - x & q_{a,c} \\ p_{b,d} - y & q_{b,d} \end{pmatrix} = x^{a+b-1}D_0 + O_x(x^{a+b-2})$$

with $D_0 \in \mathbf{Z}[y]$, with $D_0 \neq 0$ and monic, up to a sign, and $\deg(D_0) = |d-c| - 1 \geq 0$.

Proof. Working in $\tilde{R}_2$, write

$$(x,y,z) = (\delta + \delta^*, \eta + \eta^*, \delta\eta + \delta^*\eta^*).$$

It will be useful to use the notations[1] $c(\delta^n\eta^m) := \delta^n\eta^m + (\delta^*)^n(\eta^*)^m$, and $s(\delta^n\eta^m) := \delta^n\eta^m - (\delta^*)^n(\eta^*)^m$, interpreting δ^{-1} as δ^* as before to extend the definitions of c and s to include negative powers of δ and η. Note that analogs of trigonometric formulas hold also for these functions.

Now,

$$\begin{aligned} (g_*(x), g_*(y)) &= (c(\delta^a\eta^c), c(\delta^b\eta^d)) \\ &= \frac{1}{2}(c(\delta^a)c(\eta^c) + s(\delta^a)s(\eta^c), c(\delta^b)c(\eta^d) + s(\delta^b)s(\eta^d)). \end{aligned} \tag{3.2}$$

To continue our calculation we introduce *Chebyshev polynomials*. For $n \in \mathbf{N} \cup \{0\}$ the nth Chebyshev polynomial of the *first kind* is the unique element $T_n \in \mathbf{Z}[t]$ such that $T_n(\cos(\theta)) = \cos(n\theta)$ for $\theta \in [0, 2\pi]$. The polynomial $2T_n\left(\frac{t}{2}\right)$ has integer coefficients and is monic in t of degree n. The nth Chebyshev polynomial of the *second kind* is the unique element $U_n \in \mathbf{Z}[t]$ such that $U_n(\cos(\theta)) = \frac{\sin((n+1)\theta)}{\sin(\theta)}$ for $\theta \in (0, \pi)$. The polynomial $U_n\left(\frac{t}{2}\right)$ has integer coefficients and is monic in t of degree n. Because of their definition in terms of trigonometric functions, the T_n and U_n satisfy certain identities. For example, the difference of angles formula for sines gives for $m > n$

$$U_{m-n-1} = U_{m-1}T_n - U_{n-1}T_m. \tag{3.3}$$

We have

$$c(\delta^a) = 2T_a\left(\frac{x}{2}\right), \quad c(\eta^c) = 2T_c\left(\frac{y}{2}\right). \tag{3.4}$$

[1] $c(\delta)$ should be thought of as $2\cos(\theta)$ for abstract θ such that $\delta = \exp(i\theta)$.

Similarly for $a, c \geq 2$

$$s(\delta^a) = U_{a-1}\left(\frac{x}{2}\right) s(\delta), \quad s(\eta^c) = U_{c-1}\left(\frac{y}{2}\right) s(\eta). \tag{3.5}$$

Although we work in $\mathbf{Z}[\frac{1}{2}] \otimes \tilde{R}_2$ throughout the proof, our final results will hold in R_2.

We obtain from (3.2) and (3.4), (3.5) the expression

$$(g_*(x), g_*(y)) = (P_{a,c}(x,y,z), P_{b,d}(x,y,z)),$$

where

$$P_{a,c}(x,y,z) := 2T_a\left(\frac{x}{2}\right) T_c\left(\frac{y}{2}\right) + \frac{1}{2} U_{a-1}\left(\frac{x}{2}\right) U_{c-1}\left(\frac{y}{2}\right)(2z - xy).$$

To obtain this expression, we used that $s(\delta)s(\eta) = 2z - xy$. The key point is that $P_{a,c}(x,y,z)$ is linear in z, and we obtain Part 1 of the proposition with

$$\begin{aligned} p_{a,c}(x,y) &:= 2T_a\left(\frac{x}{2}\right) T_c\left(\frac{y}{2}\right) - \frac{1}{2} xy U_{a-1}\left(\frac{x}{2}\right) U_{c-1}\left(\frac{y}{2}\right), \\ q_{a,c}(x,y) &:= U_{a-1}\left(\frac{x}{2}\right) U_{c-1}\left(\frac{y}{2}\right). \end{aligned}$$

Using that $2T_a\left(\frac{t}{2}\right)$ and $U_{a-1}\left(\frac{t}{2}\right)$ are monic in t for $a \geq 1$ of degrees a and $a-1$ respectively, we get that the leading x-degree contribution to $p_{a,c}$ is $x^a u_c$ where

$$u_c(y) := T_c\left(\frac{y}{2}\right) - \frac{y}{2} U_{c-1}\left(\frac{y}{2}\right). \tag{3.6}$$

The leading x-degree contribution to $q_{a,c}$ is more easily seen to be $x^{a-1} v_c$ where

$$v_c(y) := U_{c-1}\left(\frac{y}{2}\right).$$

This concludes our calculations for the pair a, c. Since g is good, we have $b, d \geq 2$ and so the calculation of $P_{b,d}$ and $p_{b,d}$, $q_{b,d}$ is analogous to the preceding one, replacing $a, c \mapsto b, d$.

Calculation of D and D_0. Note that since g is good, we must have $c \neq d$. Indeed, if $c = d$ then $c(a-b) = \pm 1$ which cannot happen since $c \geq 2$. Since $a \geq 2$, the $-x$ term in the determinant does not contribute to the largest x-degree term. We get from (3.6)

$$\begin{aligned} D &= x^{a+b-1}\left(u_c v_d - v_c u_d\right) + O_x(x^{a+b-2}) \\ &= x^{a+b-1}\left(T_c\left(\frac{y}{2}\right) U_{d-1}\left(\frac{y}{2}\right) - T_d\left(\frac{y}{2}\right) U_{c-1}\left(\frac{y}{2}\right)\right) + O_x(x^{a+b-2}) \\ &= x^{a+b-1} \mathrm{sign}(d-c) U_{|d-c|-1}\left(\frac{y}{2}\right) + O_x(x^{a+b-2}), \end{aligned}$$

where the last equality used (3.3). □

The next step in the proof of Proposition 3.2 is to establish some general properties of the polynomials that appear in the action of $\mathrm{Out}(\mathbf{F}_2)$ on $\mathbb{X}$. By work of Fricke-Klein [9], for any $w \in \mathbf{F}_2$, the induced *word map* $w : \mathrm{SL}_2(\mathbf{C}) \times \mathrm{SL}_2(\mathbf{C}) \to \mathrm{SL}_2(\mathbf{C})$ has

$$\mathrm{tr}(w(A,B)) = P_w(x,y,z)$$

for unique $P_w \in \mathbf{Z}[x,y,z]$, where $x = \mathrm{tr}(A), y = \mathrm{tr}(B), z = \mathrm{tr}(AB)$. Indeed this follows from repeated applications of the identity (2.3). If $\theta \in \mathrm{Aut}(\mathbf{F}_2)$ acts by $\theta(X,Y) = (w_1(X,Y), w_2(X,Y))$ then θ acts on the coordinate functions $x, y \in R$ by

$$\theta_*(x) = P_{w_1}, \quad \theta_*(y) = P_{w_2}, \quad P_{w_i} \in \mathbf{Z}[x,y,z].$$

Define the (x,z)-degree of a monomial $x^\alpha y^\beta z^\gamma \kappa^\delta$ to be $\alpha+\gamma$, and define the (x,z)-degree of a polynomial f in $\mathbf{Z}[\kappa,x,y,z]$ to be the maximum of the (x,z)-degrees of the monomials with nonzero coefficients in f. We write $f^{(N)}$ for the (x,z)-degree N piece of f, that is, the part comprised of monomials of (x,z)-degree N.

Lemma 3.5. *Write X, Y for fixed generators of $\mathbf{F}_2$. Let*

$$w = X^{\alpha_1} Y^{\beta_1} X^{\alpha_2} Y^{\beta_2} \dots X^{\alpha_t} Y^{\beta_t}$$

be a monotone word, with every $\alpha_i, \beta_i \neq 0$. Let $a = \sum_{i=1}^t \alpha_i$ and $b = \sum_{i=1}^t \beta_i$. The (x,z)-degree of P_w is $\leq |a|$.

Proof. Assume for ease of exposition that all α_i, β_i are positive, so $a, b > 0$. This will be the case for words arising from good elements of $\mathrm{GL}_2(\mathbf{Z})$. The proof is by induction on the partial order $\preceq$ defined by the following moves:

- If any α_i has $\alpha_i \geq 2$ then $w', w'' \preceq w$ for either w', w'' obtained by replacing $\alpha_i \mapsto \alpha_i - 1$ or $\alpha_i \mapsto \alpha_i - 2$. Then (2.3) yields
$$P_w(x,y,z) = xP_{w'}(x,y,z) - P_{w''}(x,y,z).$$
Note if the lemma holds for $P_{w'}$ and $P_{w''}$, it holds for P_w.
- If any β_i has $\beta_i \geq 2$ then we perform the replacements $\beta_i \mapsto \beta_i - 1$ or $\beta_i \mapsto \beta_i - 2$ to form w', w'' and declare $w', w'' \preceq w$. By the same logic as before, $P_w(x,y,z) = yP_{w'}(x,y,z) - P_{w''}(x,y,z)$ so if the lemma holds for w' and w'' it holds for w.
- We identify all words with their cyclically reduced conjugates. This doesn't change P_w.

To put this all together, note that any minimal cyclically reduced word with respect to $\preceq$ has all the $\alpha_i = \beta_i = 1$. If all the α_i and β_i are 1, and w is cyclically reduced, then w is a power of XY or YX and e.g. if $w = (XY)^n$ then $a = n$. On the other hand, $P_{(XY)^n}(x,y,z) = 2T_n\left(\frac{z}{2}\right)$ has (x,z)-degree n as required (this also shows the statement of the lemma is sharp). □

Our next goal is to show, in the present context, that the P_w are equal in R to functions that are linear in z and such that certain terms have no dependence on κ.

Lemma 3.6. *If $(X,Y) \mapsto (w_1(X,Y), w_2(X,Y))$ is in $\mathrm{Aut}(\mathbf{F}_2)$, then*

$$P_{w_1}(x,y,z) = U_{w_1} + V_{w_1} z \mod I, \quad P_{w_2}(x,y,z) = U_{w_2} + V_{w_2} z \mod I$$

where $U_{w_i}, V_{w_i} \in \mathbf{Z}[\kappa,x,y]$ have the following property. If N_i is at least the (x,z)-degree of P_{w_i} then

(1) *$U_{w_i} = x^{N_i} U^0_{w_i} + O_x(x^{N_i-1})$ with $U^0_{w_i} \in \mathbf{Z}[y]$.*
(2) *$V_{w_i} = x^{N_i-1} V^0_{w_i} + O_x(x^{N_i-2})$ with $V^0_{w_i} \in \mathbf{Z}[y]$.*

In particular, $U^0_{w_i}$ and $V^0_{w_i}$ do not depend on κ.

Proof. Transform $P_{w_1}(x,y,z)$ by replacing each monomial of the form $x^\alpha y^\beta z^\gamma$ with $\gamma \geq 2$ by

$$x^\alpha y^\beta z^\gamma \mapsto x^\alpha y^\beta z^{\gamma-2}(xyz - x^2 - y^2 + 2 + \kappa), \tag{3.7}$$

these two terms are equal $\bmod I$. Moreover this replacement has the following properties: if $p, q \in \mathbf{Z}[\kappa, x, y, z]$ and $p \mapsto q$ in this manner then

- The (x,z)-degree of q is at most the (x,z)-degree of p.
- Let N_1 be at least the (x,z)-degree of p and let $p^{(N_1)}$ be the (x,z)-degree N_1 component of p and similarly define $q^{(N_1)}$. If N_1 is larger than the (x,z)-degree of p then $p^{(N_1)}$ is zero. If $p^{(N_1)} \in \mathbf{Z}[x,y,z]$ then $q^{(N_1)} \in \mathbf{Z}[x,y,z]$ (so doesn't depend on κ). This follows since $q^{(N_1)}$ is obtained from $p^{(N_1)}$ by replacement of all monomials of the form $x^\alpha y^\beta z^\gamma$ with $\gamma \geq 2$ by
$$x^\alpha y^\beta z^\gamma \mapsto x^\alpha y^\beta z^{\gamma-2}(xyz - x^2) = x^{\alpha+1} y^\beta z^{\gamma-2}(yz - x).$$
Monomials $x^\alpha y^\beta z^\gamma$ with $\gamma \leq 1$ are left unaltered.

The effect of iterating this reduction, beginning with the fact that $P_{w_1} \in \mathbf{Z}[x,y,z]$, yields polynomials $U_{w_1}, V_{w_1} \in \mathbf{Z}[\kappa, x, y]$ such that $P_{w_1} = U_{w_1} + V_{w_1} z \bmod I$, the (x,z)-degree of $U_{w_1} + V_{w_1} z$ is $\leq N_1$, and $(U_{w_1} + V_{w_1} z)^{(N_1)} \in \mathbf{Z}[x,y,z]$. This means that the x-degree of U_{w_1} is $\leq N_1$ and $U_{w_1}^{(N_1)} \in \mathbf{Z}[x,y]$. Similarly the x-degree of V_{w_1} is $\leq N_1 - 1$ and $V_{w_1}^{(N_1-1)} \in \mathbf{Z}[x,y]$. Performing this reduction also for P_{w_2} with N_2 in place of N_1 establishes the result. □

We now have everything we need to prove Proposition 3.2.

Proof of Proposition 3.2. Let $\hat{g} \in \mathrm{Aut}(\mathbf{F}_2)$ be a monotone automorphism representing g, given by Proposition 2.2. We consider $g_*(x)$, the calculation of $g_*(y)$ is similar. Let w_1 and w_2 be the monotone words appearing in the expression (2.1) for $\hat{g}$. We have $g_*(x) = P_{w_1}(x,y,z)$. This has (x,z)-degree $\leq a$ by Lemma 3.5. Note that since we know w_1 is monotone, we can conjugate w_1 to be of the form as in Lemma 3.5 without changing a or $P_{w_1}(x,y,z)$.

Applying Lemma 3.6 with $N_1 = a$ we can write

$$g_*(x) = U_{w_1}^0 x^a + U'_{w_1} + (V_{w_1}^0 x^{a-1} + V'_{w_1}) z,$$

where $U_{w_1}^0, V_{w_1}^0 \in \mathbf{Z}[y]$, $U'_{w_1} \in \mathbf{Z}[\kappa, x, y]$ has x-degree $\leq a-1$ and $V'_{w_1} \in \mathbf{Z}[\kappa, x, y]$ has x-degree $\leq a-2$. We obtain the first part of the proposition with

$$\tilde{p}_{a,c} := U_{w_1}^0 x^a + U'_{w_1}, \quad \tilde{q}_{a,c} := V_{w_1}^0 x^{a-1} + V'_{w_1}. \tag{3.8}$$

Similarly $\tilde{p}_{c,d}$ and $\tilde{q}_{c,d}$ are obtained by replacing w_1 by w_2 and $a, c \mapsto b, d$. Note that at this moment we do not know that $U_{w_1}^0$ and $U_{w_2}^0$ are non-zero.

Let π be the evaluation map $\mathbf{Z}[\kappa, x, y, z] \to \mathbf{Z}[x, y, z]$ sending $\kappa \mapsto 2$. We must have in R_2

$$\pi\left(\tilde{p}_{a,c} + \tilde{q}_{a,c} z\right) \equiv \pi(\tilde{p}_{a,c}) + \pi(\tilde{q}_{a,c}) z \equiv p_{a,c} + q_{a,c} z \bmod I_2 \tag{3.9}$$

where $p_{a,b}$ and $q_{a,b}$ are the polynomials from Proposition 3.4. This is because they both describe how g maps the coordinate function x. In the other hand, since the left and right hand sides of (3.9) differ by a function that is linear in z, this difference

must be zero since 0 is the only element of I_2 that is linear in z. So the identity (3.9) actually holds in $\mathbf{Z}[x,y,z]$. This means $\pi(\tilde{p}_{a,c}) = p_{a,c}$, $\pi(\tilde{q}_{a,c}) = \pi(q_{a,c})$ and the same replacing $a, c \mapsto b, d$.

This implies, if D is the quantity obtained in Proposition 3.4, that

$$\pi(\tilde{D}) = D.$$

From (3.8) we have

$$\tilde{D} = x^{a+b-1}(U^0_{w_1}V^0_{w_2} - U^0_{w_2}V^0_{w_1}) + O_x(x^{a+b-2}).$$

Since the x^{a+b-1} coefficient of $\tilde{D}$ doesn't depend on κ, and equals D_0 (from Proposition 3.4) when evaluated at $\kappa = 2$, it must be equal to this D_0. This completes the proof, using the properties of D_0 described in Proposition 3.4. □

4. Algebraic Characterization of Cyclic Palindromes

In this section, we use that $\mathrm{PSL}_2(\mathbf{Z}) \cong \mathbf{Z}/2\mathbf{Z} * \mathbf{Z}/3\mathbf{Z}$ with the generators of the cyclic factors given by

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}.$$

Here $R = ST$ where $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. With this presentation, every conjugacy class in S has a representative of the form either $g = R^y$, $g = S$, or

$$g = SR^{y_1} \dots SR^{y_k} \tag{4.1}$$

with $y_i \in \{1,2\}$ for $1 \le i \le k$. However, powers of S and R are not hyperbolic, so every hyperbolic conjugacy class has a representative as in (4.1). Moreover, a representative of this form has unique sequence $y_1, \dots, y_k$, up to cyclic rotation. We write $[y_1, \dots, y_k]$ for the cyclic equivalence class of this sequence.

Proof of Lemma 1.6. Note that in $\mathrm{PSL}_2(\mathbf{Z})$,

$$SR = \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix} = V, \quad SR^2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = U, \tag{4.2}$$

and substituting this into (4.1) proves Lemma 1.6. □

Proof of Proposition 1.9. Suppose $g = U^{n_1}V^{m_1} \dots U^{n_l}V^{m_l}$ is a hyperbolic reduced UV-word. Also suppose g is given by (4.1). Then g^{-1} is conjugate in $\mathrm{PSL}_2(\mathbf{Z})$ to

$$SR^{(1-y_k)}SR^{(1-y_{k-1})} \dots SR^{(1-y_1)}. \tag{4.3}$$

The action of $\mathrm{PGL}_2(\mathbf{Z})$ on conjugacy classes in $\mathrm{PSL}_2(\mathbf{Z})$ is generated by

$$w = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We calculate

$$wSw^{-1} = S, \quad wRw^{-1} = R^2.$$

Therefore with g as in (4.1), we have

$$wgw^{-1} = SR^{(1-y_1)}S \dots SR^{(1-y_k)}. \tag{4.4}$$

which is conjugate in $\mathrm{PSL}_2(\mathbf{Z})$ to g^{-1}. Then comparing (4.3) and (4.4) we have that g is ambiguous if and only if $[(1-y_1),(1-y_2),\dots,(1-y_k)] = [(1-y_k),(1-y_{k-1})\dots(1-y_1)]$ which is if and only if $[y_1,y_2,\dots,y_k] = [y_k,\dots,y_1]$, and it is easy to see, using the substitutions (4.2), that this happens if and only if the reduced UV-word giving g is a cyclic palindrome. □

5. Proof of Theorem 1.2

In this section we prove Theorem 1.2. We do this by calculating the sign of the Markoff moves m_i and elements of S_3 as permutations of $\mathbb{Y}_{-2}(\mathbb{F}_p)$.

We begin by examining the subgroup N of $\mathrm{Aut}(\mathbb{X}_\kappa)$ as it plays a special role in the action of $\mathrm{Out}(\mathbf{F}_2)$ on $\mathbb{X}^*_{-2}(\mathbb{F}_p)$. Recall from the Introduction the elements n_i and the fact that $\mathrm{Out}(\mathbf{F}_2)$ permutes the N-orbits of $\mathbb{X}^*_{-2}(\mathbb{F}_p)$.

Lemma 5.1. *There are no points in $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ with zeroes in exactly two coordinate entries. Hence for $p>2$ all orbits of N in $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ are of size 4.*

Proof. By symmetry, it suffices to check that we can have no $(0,0,z)\in\mathbb{X}^*_{-2}(\mathbb{F}_p)$, with $z\neq 0$. If $x,y=0$, substituting into (1.1) with $\kappa=-2$ we obtain $0+0+z^2=0$ which implies $z=0$. Given the first statement of the lemma, the second follows since no points of $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ are fixed by any n_i. □

Due to a result of Carlitz [5], $|\mathbb{X}^*_{-2}(\mathbb{F}_p)| = p(p+3)$ when $p\equiv 1 \pmod 4$ and $|\mathbb{X}^*_{-2}(\mathbb{F}_p)| = p(p-3)$ when $p\equiv 3 \pmod 4$. Thus,

$$|\mathbb{Y}_{-2}(\mathbb{F}_p)| = \begin{cases} \frac{1}{4}p(p+3), & \text{if } p\equiv 1 \pmod 4 \\ \frac{1}{4}p(p-3), & \text{if } p\equiv 3 \pmod 4 \end{cases} \tag{5.1}$$

The following fact will be useful later.

Fact 5.2. *In $\mathbb{F}_p$, the number of distinct pairs of consecutive quadratic residues, both nonzero, is exactly:*

$$\begin{cases} \frac{1}{4}(p-5), & \textit{when } p\equiv 1 \pmod 4 \\ \frac{1}{4}(p-3), & \textit{when } p\equiv 3 \pmod 4 \,. \end{cases} \tag{5.2}$$

The total number of consecutive quadratic residues is found in [1, Theorem 10-2][2]. We discount the pair $(0,1)$ in both cases, and $(-1,0)$ when $p\equiv 1 \pmod 4$.

Lemma 5.3. *Let p be an odd prime. For a given $i\in\{1,2,3\}$*

$$\#\{(x,y,z)\in\mathbb{X}^*_{-2}(\mathbb{F}_p) \mid m_i(x,y,z)=(x,y,z)\} = \begin{cases} p-5, & p\equiv 1 \pmod 4 \\ p-3, & p\equiv 3 \pmod 4 \,. \end{cases}$$

[2]Count the number of the solutions (a,b) to $a^2-b^2=1$ in $\mathbb{F}_p$. To do this, count unordered pairs $\alpha := a+b,\ \beta := a-b$ such that $\alpha\beta=1$, then discount ones that result in the same values of a^2, b^2.

Proof. We will prove this formula for m_1, and it follows for m_2, m_3 by symmetry. We have that $m_1(x,y,z) = (x,y,z)$ exactly when

$$2x = yz. \tag{5.3}$$

Lemma 5.1, equation (5.3), and our assumption that $(x,y,z) \neq (0,0,0)$ imply that $x, y, z \neq 0$. Substituting $x = yz/2$ into (1.1) we have

$$y^2 + z^2 - \frac{y^2 z^2}{4} = 0. \tag{5.4}$$

As x is uniquely determined given y, z by (5.3) we count the solutions to (5.4) over $\mathbb{F}_p$.

Letting $Y = y^2, Z = z^2$ we have

$$Z(Y-4) = 4Y\,. \tag{5.5}$$

As $y, z \neq 0$ there are exactly as many y, z satisfying (5.4) as four times the number of solutions to (5.5).

By (5.5), as $Y \neq 0$, Z is determined uniquely by Y, so we just need to count possible values of $Y \neq 0$ that can satisfy (5.5). As Y and Z are quadratic residues, $Y - 4$ must also be. Thus to count the possible solutions to (5.5), we just need to count the possible values of Y such that both Y and $Y - 4$ are nonzero quadratic residues. This is the case if and only if $Y/4$ and $(Y-4)/4$ are consecutive nonzero quadratic residues. By (5.2), for $p \equiv 1 \pmod 4$ (resp. $p \equiv 3 \pmod 4$), there are $(p-5)/4$ (resp. $(p-3)/4$) of these. This gives us our result. □

Lemma 5.4. *Suppose p is an odd prime. For a given $i \in \{1,2,3\}$, the Markoff move m_i acts as an even permutation on $\mathbb{Y}_{-2}(\mathbb{F}_p)$ exactly when $p \equiv 3 \pmod 8$.*

Proof. We will show this result for m_1 and it follows by symmetry for m_2, m_3. Because it is an involution, the permutation induced by m_1 on $\mathbb{Y}_{-2}(\mathbb{F}_p)$ is a product of

$$r := \frac{|\mathbb{Y}_{-2}(\mathbb{F}_p)| - |F|}{2} \tag{5.6}$$

disjoint transpositions, where F is the set of fixed points of m_1 in $\mathbb{Y}_{-2}(\mathbb{F}_p)$. Each of the n_i commute with m_1, so $\mathbf{x} \in \mathbb{X}_{-2}(\mathbb{F}_p)$ is fixed by m_1 if and only if all the elements of $N \cdot \mathbf{x}$ are fixed by m_1. Consequently $|F|$ is exactly one fourth the number of fixed points of m_1 in $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ which we have calculated in Lemma 5.3. We also recall from (5.1) the size of $\mathbb{Y}_{-2}(\mathbb{F}_p)$. We calculate the parity of m_1 by calculating r case by case:

If $p = 4k+1$

$$r = \frac{1}{2}\left(\frac{p^2+3p}{4} - \frac{p-5}{4}\right) = 2(k^2+k)+1 \equiv 1 \pmod 2,$$

so m_1 acts as an odd permutation. If $p = 8k+7$

$$r = \frac{1}{2}\left(\frac{p^2-3p}{4} - \frac{p-3}{4}\right) = 8k^2+10k+3 \equiv 1 \pmod 2,$$

so m_1 acts as an odd permutation. If $p = 8k + 3$

$$r = \frac{1}{2}\left(\frac{p^2 - 3p}{4} - \frac{p-3}{4}\right) = 8k^2 + 2k \equiv 0 \pmod 2,$$

so m_1 acts as an even permutation. □

Proposition 5.5. *The permutation group generated by the action of $\langle m_1, m_2, m_3\rangle$ on $\mathbb{Y}_{-2}(\mathbb{F}_p)$ is contained in the alternating group on $\mathbb{Y}_{-2}(\mathbb{F}_p)$ if and only if $p \equiv 3 \pmod 8$.*

Proof. This follows directly from Lemma 5.4. □

In order to complete our proof of Theorem 1.2, we must check the parity of the other generators of $\mathrm{PGL}_2(\mathbf{Z})$ (through which $\mathrm{Out}(\mathbf{F}_2)$ acts). The only remaining generators to check, aside from the Markoff moves, are those of S_3. By Proposition 5.5, we know there always will be odd permutations for $p \not\equiv 3 \pmod 8$, so we only need to examine the remaining case, when $p \equiv 3 \pmod 8$.

Lemma 5.6. *The action of S_3 on $\mathbb{Y}_{-2}(\mathbb{F}_p)$ consists of even permutations when $p \equiv 3 \pmod{16}$. When $p \equiv 11 \pmod{16}$, it consists of both even and odd permutations.*

Proof. The group S_3 is generated by transpositions, and by symmetry they all have the same parity, so it suffices to check the parity of the action of the transposition $(1\,2)$ in the cases we consider.

Our strategy is to count the points in $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ whose N-orbits are fixed by the transposition $(1\,2)$. We start by counting how many possible values x can take on, then for each of those values we will count how many points with fixed orbits there are.

The N-orbit of (x, y, z) is fixed by $(1\,2)$ if and only if

$$(x, y, z) \in \{(y, x, z), (y, -x, -z), (-y, x, -z), (-y, -x, z)\}, \tag{5.7}$$

which is if and only if $x = \pm y$. Note that by Lemma 5.1 this rules out $x = 0$.

Substituting $x = \pm y$ into (1.1) with $\kappa = -2$ we reduce to two cases:

$$x \neq 0, \quad y = x, \quad 2x^2 + z^2 = x^2 z, \text{ or} \tag{5.8}$$

$$x \neq 0, \quad y = -x, \quad 2x^2 + z^2 = -x^2 z\ . \tag{5.9}$$

For fixed x, in both cases we obtain quadratic equations in z with discriminant $\Delta = x^2(x^2 - 8)$. Note that $\Delta \neq 0$ as $x \neq 0$ and 8 is not a quadratic residue of $\mathbb{F}_p$ because $p \equiv 3 \pmod 8$ in the cases we consider. Thus (5.8) and (5.9) have solutions over $\mathbb{F}_p$ if and only if Δ is a square, which happens if and only if $x^2 - 8$ is a square.

As we assume $p \equiv 3 \pmod 8$, there exists[3] some α such that $\alpha^2 = -8$. Setting $w := x/\alpha$ we want to count how many values w can take such that $x^2 - 8 = -8(w^2 + 1)$ is a square, which we do by counting the number of nonzero consecutive quadratic residues w^2 and $w^2 + 1$. From Fact 5.2 we have that there are $(p-3)/4$

[3]As $p \equiv 3 \pmod 4$ we have that $\left(\frac{-1}{p}\right) = -1$ and as $p \equiv 3 \pmod 8$ we have that $\left(\frac{2}{p}\right) = -1$. This implies that $\left(\frac{-2}{p}\right) = \left(\frac{-8}{p}\right) = 1$.

such pairs of the form (w^2, w^2+1) where $w^2 \neq 0$ (as in both cases $p \equiv 3 \pmod 4$). Each pair of residues, (w^2, w^2+1), can be made by both w and $-w$, which gives us $(p-3)/2$ possible values of w and hence of x.

For each valid x, those such that Δ is a square, we have exactly four solutions total to (5.8) and (5.9) for (x, y, z) that correspond to four points which satisfy both (1.1) and (5.7) and thus four points whose N-orbits are fixed by $(1\,2)$:

$$(x, x, z_1), (x, x, z_2), (x, -x, -z_1), (x, -x, -z_2)$$
$$\text{where } z_1 = \frac{x^2 + \sqrt{\Delta}}{2},\ z_2 = \frac{x^2 - \sqrt{\Delta}}{2}.$$

Recall that as $\Delta \neq 0$, we have that $z_1 \neq z_2$, so these four points are distinct. This gives us $2(p-3)$ points of $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ in total whose N-orbits are fixed by $(1\,2)$. As each N-orbit in $\mathbb{X}^*_{-2}(\mathbb{F}_p)$ has exactly 4 points, there are $\frac{p-3}{2}$ fixed N-orbits of $(1\,2)$.

To determine the parity of $(1\,2)$, we use the same method of counting disjoint transpositions as we did for m_1 in the proof of Lemma 5.4. Letting F denote the fixed N-orbits of $(1\,2)$, we examine the two cases:

If $p = 16k + 3$

$$\frac{|\mathbb{Y}_{-2}(\mathbb{F}_p)| - |F|}{2} = \frac{1}{2}\left(\frac{p(p-3)}{4} - \frac{p-3}{2}\right) = 2k(16k+1) \equiv 0 \pmod 2,$$

so $(1\,2)$ acts as an even permutation.

If $p = 16k + 11$

$$\frac{|\mathbb{Y}_{-2}(\mathbb{F}_p)| - |F|}{2} = \frac{1}{2}\left(\frac{p(p-3)}{4} - \frac{p-3}{2}\right) = 32k^2 + 34k + 9 \equiv 1 \pmod 2,$$

so $(1\,2)$ acts as an odd permutation. The lemma follows directly from this result. □

Theorem 1.2 now follows directly from Lemma 5.6 and Proposition 5.5.

References

[1] G. E. Andrews. *Number theory*. Dover Publications, Inc., New York, 1994. Corrected reprint of the 1971 original [Dover, New York; MR0309838 (46 #8943)].

[2] J. Bourgain, A. Gamburd, and P. Sarnak. Markoff Surfaces and Strong Approximation: 1. *arXiv:1607.01530*, July 2016.

[3] B. H. Bowditch. Markoff triples and quasifuchsian groups. *Proceedings of the London Mathematical Society*, 77(3):697–736, 1998.

[4] S. Cantat. Bers and Hénon, Painlevé and Schrödinger. *Duke Math. J.*, 149(3):411–460, 2009.

[5] L. Carlitz. The number of solutions of some special equations in a finite field. *Pacific Journal of Mathematics*, 4:207–217, 1954.

[6] A. Cayley. A memoir on cubic surfaces. *Philosophical Transactions of the Royal Society of London*, 159:231–326, 1869.

[7] F. X Connolly and J. F Davis. The surgery obstruction groups of the infinite dihedral group. *Geom. Topol.*, 8(3):1043–1078, 2004.

[8] R. Fricke. Über die Theorie der automorphen Modulgrupper. *Nachr. Akad. Wiss. Göttingen* 91–101, 1896

[9] R. Fricke and F. Klein. *Vorlesungen über die Theorie der automorphen Funktionen. Band 1: Die gruppentheoretischen Grundlagen. Band II: Die funktionentheoretischen Ausführungen und die Andwendungen*, volume 4 of *Bibliotheca Mathematica Teubneriana, Bände 3*. Johnson Reprint Corp., New York; B. G. Teubner Verlagsgesellschaft, Stuttg art, 1965.

[10] A. Gál and P. B. Miltersen. The cell probe complexity of succinct data structures. *Theoretical computer science*, 379(3):405–417, 2007.
[11] E. Ghys and V. Sergiescu. Stabilite et conjugaison differentiable pour certains feuilletages. *Topology*, 19(2):179 – 197, 1980.
[12] W. Goldman. The modular group action on real $SL(2)$-characters of a one-holed torus. *Geometry & Topology*, 7:443–486, 2003.
[13] D. A. Hejhal. *The Selberg Trace Formula for* $\mathrm{PSL}(2,\mathbf{R})$*, Volume 2.* Springer-Verlag Berlin Heidelberg, 1983.
[14] R. Horowitz. Induced automorphisms on Fricke characters of free groups. *Transactions of the American Mathematical Society*, 208:41–50, 1975.
[15] M. Kohmoto, L. P. Kadanoff, and C. Tang. Localization problem in one dimension: mapping and escape. *Phys. Rev. Lett.*, 50(23):1870–1872, 1983.
[16] S. Lang and A. Weil. Number of points of varieties in finite fields. *American Journal of Mathematics*, 76(4):819–827, 1954.
[17] A. Markoff. Sur les formes quadratiques binaires indéfinies. *Mathematische Annalen*, 15:381–406, 1880.
[18] D. McCullough and M. Wanderley. Nielsen equivalence of generating pairs of $\mathrm{SL}(2,q)$. *Glasg. Math. J.*, 55(3):481–509, 2013.
[19] C. Meiri, D. Puder, and D. Carmon. The Markoff Group of Transformations in Prime and Composite Moduli. *Duke Math. J.* Volume 167, Number 14, 2679-2720, 2018.
[20] J. Nielsen. Die Isomorphismen der allgemeinen unendlichen Gruppe mit zwei Erzeugenden. *Math. Ann.*, 71 385–397, 1918.
[21] O. Parzanchevski and D. Puder. Stallings graphs, algebraic extensions and primitive elements in F_2. *Math. Proc. Cambridge Philos. Soc.*, 157(1):1–11, 2014.
[22] L. Polterovich and Z. Rudnick. Kick stability in groups and dynamical systems. *Nonlinearity*, 14(5):1331, 2001.
[23] P. Sarnak. Reciprocal geodesics. *Clay Math Proceedings*, 7:217–237, 2007.
[24] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. Assoc. Comput. Mach.*, 27(4):701–717, 1980.
[25] A. Selberg. Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *J. Indian Math. Soc. (N.S.)*, 20:47–87, 1956.
[26] J. H. Silverman. Variation of periods modulo p in arithmetic dynamics. *New York Journal of Mathematics*, 14, 07 2007.
[27] W. P. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)*, 19(2):417–431, 1988.
[28] R. Zippel. Probabilistic algorithms for sparse polynomials. In *Symbolic and algebraic computation (EUROSAM '79, Internat. Sympos., Marseille, 1979)*, volume 72 of *Lecture Notes in Comput. Sci.*, pages 216–226. Springer, Berlin-New York, 1979.

Alois Cerbu,
Department of Mathematics,
University of California, Berkeley,
970 Evans Hall #3840,
Berkeley, CA 94720-3840 USA
cerbu@berkeley.edu

Michael Magee,
Department of Mathematical Sciences,
Durham University,
Lower Mountjoy, Stockton Rd,
Durham DH1 3LE, U.K.
michael.r.magee@durham.ac.uk

Elijah Gunther,
Department of Mathematics,
University of Pennsylvania,
David Rittenhouse Lab,
209 South 33rd Street,
Philadelphia, PA 19104-6395 USA
elijahg@math.upenn.edu

Luke Peilen,
Courant Institute
of Mathematical Sciences,
New York University,
251 Mercer St #801,
New York, NY 10012, USA
lkp279@nyu.edu